Supplementary material

I) Data extraction and organisation

II) Study design

III) Additional analyses & methods

I) Data extraction and organisation

This study is based on a large data set linking public health care registers in Norway to demographic information about the patients and characteristics of their primary care physicians. It comprises all contacts with the primary care and specialised services in Norway from 2008-2016, except contacts with, and admissions to psychiatric hospitals. The data of this study are publicly available, but restrictions apply to the availability. These data were used under license for the current study.

All patient contacts with Norwegian publicly funded general practitioner services generate a unique claim for reimbursement. These claims are submitted from the treating physician or service to the Norwegian Health Economics Administration and are registered in the Control and Payment of Health Reimbursement Register (KUHR). (1) For this study, we used the reimbursement claims data from all out-of-hours contacts in Norway from 2008-16, performed between 16:00 and 07:59 on weekdays, or during both day or night on Saturday, Sunday and public holidays. These claims include patient ID, time, patient diagnoses, a unique physician identification number, and the type of contact, (e.g., telephone, consultation, home-visits, including a specific code for out-of-hours work.) Claims missing the unique physician identification number were excluded. In some out-of-hours services, physicians have a fixed salary agreement, (mostly for working night-time, but some places also for late-shifts). In these cases, the claims are filed from the service or municipality, and will thus not include the unique physician ID. The share of claims filed in such ways, vary from year to year in our material. We further selected all claims containing the specific codes for out-of-hours work where physicians assessed patients, face-to-face or by telephone (codes: 2ak, 2fk, 1ak, 1bk, 1g). We excluded claims containing codes for home visits (codes: 2nk, 11ak, 11nk, 21k). We also excluded contacts where the patient's regular GP were

present at the out-of-hours services since such contacts could have been arranged between the patient and the GP. For this study we chose to study patients who were supposedly unfamiliar to the out-of-hours physician and the staff. Therefore, we excluded claims where the patient met his or her regular GP, and where the patient had been assesses by the same physician earlier in the study period. Further, we excluded patient contacts from years where patients were frequent attenders, i.e., where they had more than four out-of-hours contacts per year (the 90-percentile for yearly contacts).

II) Study design

The programming code can be available upon request to the corresponding author.
In this study our main aim was to estimate the effects of being referred to the hospital from out-of-hours services for patients where the indications for such referral were unclear. Existing clinical guidelines will to some extent aid the referral decision, however, in many situations patients will not be easily classified in the guidelines. This could be patients with multimorbidity (2) or symptoms compatible with severe and not so severe conditions. We believe that such patients are more likely to have their referral decisions affected or determined by their physicians' referral preferences. Further, we chose to concentrate on the effects for older patients. These patients are likely more challenging to assess and highly affected by a referral decision, as they often are frailer. Using the physician preferences as an instrument in our instrumental variable model, the effects would be valid for the group of patients who have their referral decision affected by the physician preferences, hence not for patients with clear cut referral indications such as obvious myocardial infarctions or severe fractures.

To increase the contrast to the regular GP setting, we chose to only include patients who are presumably unknown to the physician and staff, as described above. We intended to investigate the effects on both patient outcomes and the dynamics of health services. Thus, our outcomes of interest were defined as health service use and mortality following the referral, both short-term and long-term. Based on this, we defined the follow-up period to be 0-10, 0-30, 0-90 and 180 days.

This study design is based on the assumption that there are no systematic associations between the treating physician's referral preference and the potentially confounding patient factors in each index contact. By including only older patients, assessed by the particular out-of-hours physician for the first time, and excluding contacts from years where patients had >4 contacts, the patients were more similar, thus decreasing the risk of such systematic differences. An example of an association we are trying to avoid would be that a physician with high referral preference would be more likely to meet patients with more severe conditions. Out-of-hours services provide a good study setting since patients usually do not choose when to get acutely ill, and we can assume that in most cases, they have no knowledge of which physicians are on-call in their area. There is often only one physician serving an area per time in the Norwegian out-of-hours services; however, there is a trend towards centralizing the out-of-hours services, resulting in larger out-of-hours services, with several physicians on-call at the same time.

Even if systematic differences in patient and physician characteristics are less likely in the out-of-hours setting than in the regular GP setting, there are still some possible confounding factors that need to be addressed:

1) Local differences between the out-of-hours services, adaptation to local conditions

We assume that there are substantial differences between the various out-of-hours services across Norway. These are most likely based on local adaptations to the patient population's needs, and to the other parts of the services, such as distance to hospital, and transportation resources. Hence, the provision of out-of-hours service in a city differs from the out-of-hours services in a rural municipality. Based on this knowledge, we only compare contacts within the same services. In Norway, all municipalities have a unique municipality code. The extensive collaboration between municipalities leads to shifting geographical locations of many out-of-hours services, especially in the scarcely populated municipalities. The KUHR claims lack information about the actual geographical location where the contact took place. Although some patients fall ill when travelling, we assume that patients visit the out-of-hours services in their home area in most of the cases. However, the physicians will mainly work within only one out-of-hours service at the time, although the same out-of-hours service may often serve the population from several municipalities. We handled this

problem by linking the unique patient ID from each claim to the municipality code where patients were registered as residents (Statistics Norway). To define the location of the out-of-hours services, we used the municipality codes´ modal value among patients within physicians per week. As a result, the municipality where most of the patients (within one week, seeing the same physician) live was defined as the municipality where the contacts for this physician took place that particular week. Thus, this could change through the time periods in the data material. If the value was the same for two or more municipalities, we chose the lower code, as this often represents a larger municipality.

2) Possible time trends over the years

To handle possible changes in both organisational factors and referral practices over time, we only compare patients within the same year.

3) Systematic differences between day shifts, late shifts and night shifts

Based on the assumption that there can be systematic differences between daytime, afternoon and nighttime in patients contacts and physicians staffing the out-of-hours-services, we chose to divide 24 hours into three different time units; late shift (16:00-23:59); night shift (00:00-07:59); day shifts (08:00-15:59) during weekends or holidays. Thus, we only compare patients contacting the out-of-hours service in the same out-of-hours shift during the day.

Since our instrument was based on the patients' sex, we compared only female patients with other female patients, and male patients with other male patients.

Based on the assumptions above, we matched patients in groups defined by combining information on patients of the same sex visiting the same out-of-hours station in the same out-of-hours shifts within the same year. For example, we compared all female patients visiting the same out-of-hours station all late shifts (16:00-23:59) in 2015. By analysing only within-group variability, we effectively controlled for all confounding that was constant within each group. To avoid the effect of possible patient selection in situations where two or more physicians were on-call at the same time, we used the weighted average of physician referral preferences within each out-of-hours shift in each service.

III Additional analyses

Table S1 Descriptive statistics comparing the study population and the whole population

Table S1. Out-of-hours contacts in Norway 2008–2016 for patients >64 years:
Characteristics of the study population and the whole population weighted by the number of index contacts.

	Study population	Whole population
All contacts	944,512	1,798,169
Mean age, years (SD)	77.4 (8.4)	77.6 (8.4)
Male (%)	404,376 (42.8)	770,635 (42.9)
Low education (%)	387,598 (41.3)	770,736 (43.1)
Immigration status (%)*	39,021 (4.1)	74,478 (4.1)
Charlson Comorbidity Index based on last hospital visit (>1)	88,976 (9,6)	200,682 (11.2)
Previous health service use		
Unplanned admission to hospital previous month (%)	90,448 (9.6)	218,363 (12.4)
Elective contact with hospital previous month (%)	193,869 (20.5)	387,109 (21.5)
Outpatient specialist contacts previous month (planned and unplanned) (%)	196,695 (20.8)	395,752 (21.0)
Primary care physician visits previous month (%)	54,321 (5.8)	200,158 (11.1)
Discharge diagnosis, last hospital contact ICD-10 Chapter I Circulatory system (%)	101,565 (10.8)	201,542 (11.2)
Discharge diagnosis, last hospital contact ICD-10 Chapter C Neoplasms (%)	48,537 (5.1)	93,034 (5.2)
Referral diagnosis group from index contact		
ICPC-2 Chapter A General and Unspecified (%)	163,450 (17.7)	355,262 (19.8)
ICPC-2 Chapter D Digestive (%)	91,721 (9.9)	182,466 (10.2)
ICPC-2 Chapter K Cardiovascular (%)	95,745 (10.4)	185,907 (10.3)
ICPC-2 Chapter L Musculoskeletal (%)	158,231 (17.2)	283,582 (15.8)
ICPC-2 Chapter R Respiratory (%)	154,982 (16.8)	297,548 (16.2)

* Completed less than 13 years of school
** Immigrants or Norwegian-born to immigrant parents
*** Standard deviation

Table S2 Balance test of confounders

Table S2 Balance test of confounders. Regression analysis showing the association between potential confounders (patients characteristics) and the exposures (physicians' referral preference in instrumental variable analyses, and acute referral in multivariable adjusted analyses).

n = 922,796	Instrumental variable analyses* (unscaled)			Instrumental variable analyses* (scaled)**			Multivariable multiadjusted analyses***		
	Beta	**LCI**	**UCI**	**Beta**	**LCI**	**UCI**	**Beta**	**LCI**	**UCI**
Age in years	0,21	-0,15	0,57	0,46	-0,34	1,26	**1,49**	**1,45**	**1,53**
Low education****	-0,01	-0,02	0,00	-0,02	-0,05	0,01	**0,01**	**0,00**	**0,01**
Immigration status*****	0,00	0,00	0,01	0,01	0,00	0,02	**0,00**	**0,00**	**0,00**
ICD-10 Chapter I	0,00	-0,01	0,01	0,01	-0,01	0,02	**0,01**	**0,01**	**0,01**
ICD-10 Chapter C	0,00	-0,01	0,00	0,00	-0,02	0,01	**0,01**	**0,01**	**0,01**
Charlson Comorbidity Index > 1	-0,01	-0,02	0,00	-0,01	-0,03	0,01	**0,03**	**0,03**	**0,03**
Health care use previous month:									
Elective hospital admission	0,01	0,00	0,02	0,02	-0,01	0,04	**0,02**	**0,02**	**0,02**
Unplanned hospital admission	0,00	-0,01	0,01	-0,01	-0,03	0,01	**0,03**	**0,03**	**0,03**
General practitioner visit	0,00	-0,01	0,01	0,00	-0,02	0,01	**0,00**	**0,00**	**0,00**
Outpatient specialist contacts	0,00	-0,01	0,01	0,01	-0,02	0,03	**0,02**	**0,02**	**0,02**

*** Adjusted for patient age, and age2, month and weekday.**

**** Scaled according to the strength of the instrument-exposure association.**

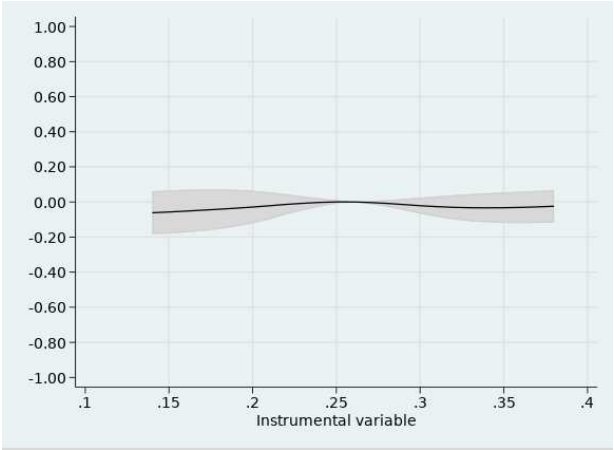
***** Adjusted for patient age, and age2, year, month and weekday and hour.**

****** Completed less than 13 years in school**

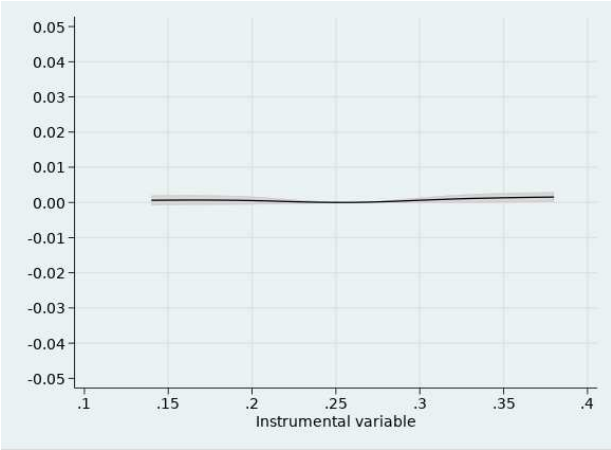
******* Immigrants and Norwegian-born to immigrant parents.**

Figure S1 Balance test of confounders, splines showing confounder associations across different values of the instrument

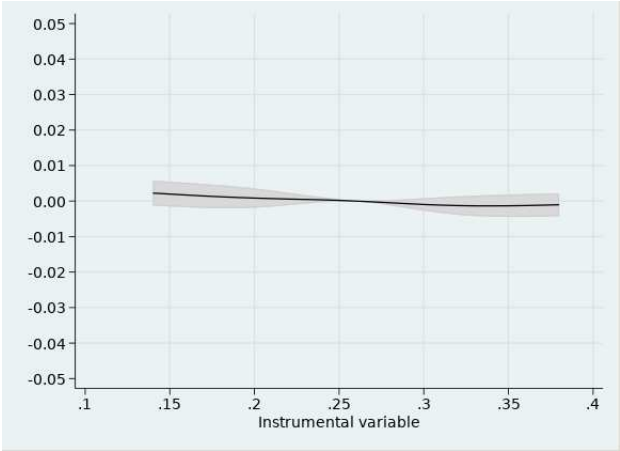
AGE



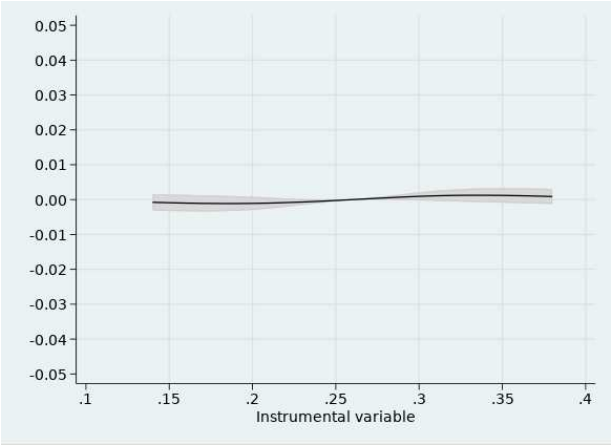
IMMIGRATION STATUS



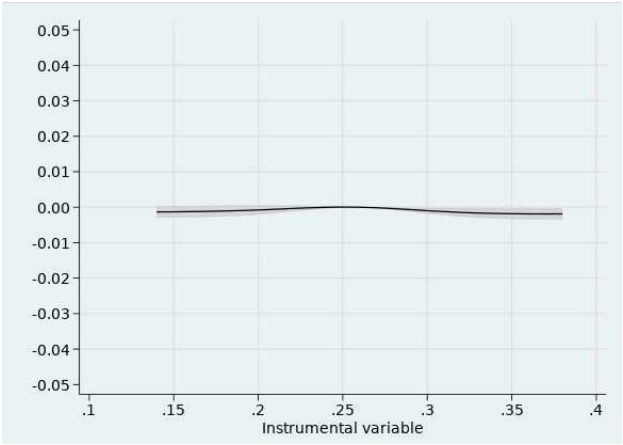
LOW EDUCATION



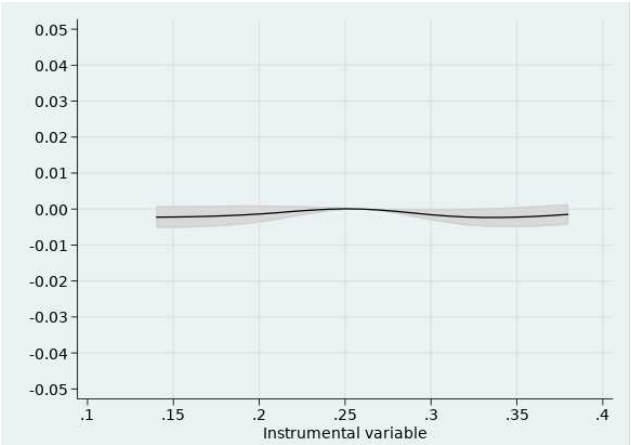
ICD10-I



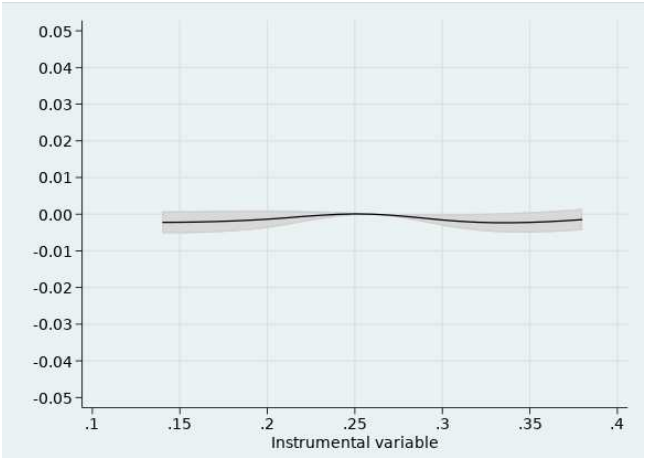
ICD10-C



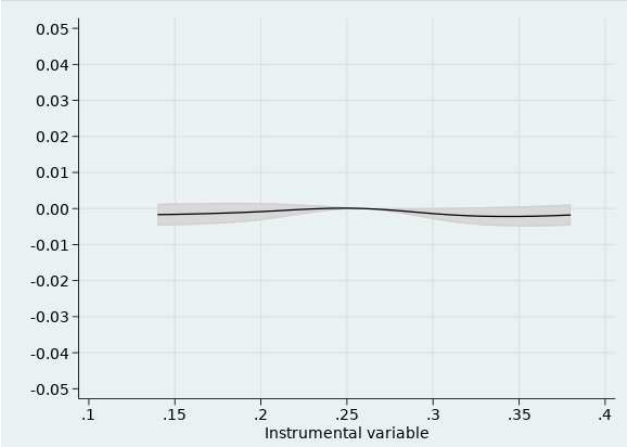
UNPLANNED HOSPITAL ADMISSION LAST 30 DAYS



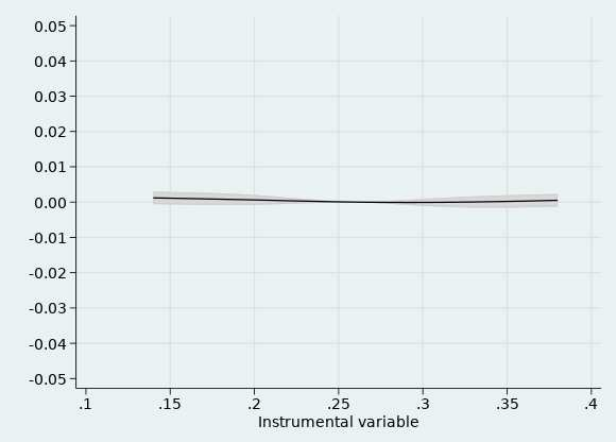
ELECTIVE ADMISSION LAST 30 DAYS



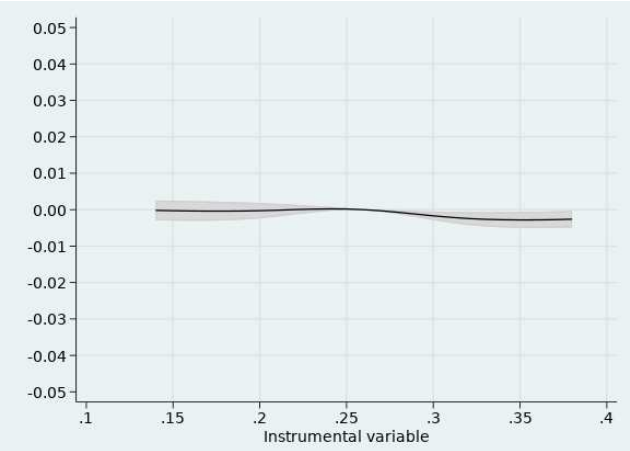
PRIMARY CARE PHYSICIAN CONTACTS LAST 30 DAYS



OUTPATIENT SPECIALIST CONTACTS LAST 30 DAYS



CHARLSON COMORBIDITY INDEX > 1



Balance tests/confounder analyses

To justify our assumption of no association between the physician referral preference and the patient characteristics for the index contacts within our defined groups, we performed balance tests where we calculated the associations between selected patient characteristics known as potential confounders, and the physician referral preference. These are shown in Supplementary Table 1. We made variables for the health service use prior to the out-of-hours contact for all patients. We also made variables based on being discharged with specific diagnoses; 1) ICD-10 Chapter IX, Diseases of the circulatory system and 2) ICD-10 Chapter III, Neoplasms (Malignant diseases). Together with variables for patient sex, age, education level and immigration status, these variables were used in analyses to test whether patient characteristics affected which GP they saw when visiting the out-of-hours services. As we can see from the results presented in Supplementary Table 2 and Supplementary Figure 1, there was little or no evidence of any differences between GP characteristics and possible patient confounders, neither across different values of the instrument, given our design. Hence, these results provide a strong support of the confounder independence assumption of our design.

Figure S2 Relevance criterium, spline showing the linear association with chance of referral across different values of the instrument.

One standard deviation (0.1) increase in the physician referral preference (instrument) was associated with a risk difference of about four percentage points for referral to the hospital.

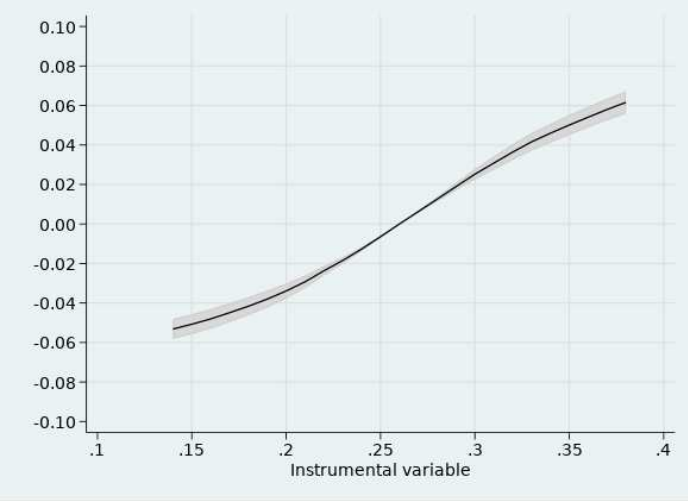


Table S3 Sensitivity analyses with different adjustments and definitions of the instrument

Table S3 Instrumental variable analyses, sensitivity analyses with different adjustments and definitions of the instrument. Mean difference in days with contacts.													
	Instrument 1*					Instrument 1, adjusted for referral diagnosis**				Instrument 2***			
Primary care physician	**Mean difference**	**LCI**	**UCI**	**F-statistics**	**P- value endogeneity test******	**Mean difference**	**LCI**	**UCI**	**F-statistics**	**Mean difference**	**LCI**	**UCI**	**F-statistics**
Within 10 days	**-0.14**	-0.21	-0.07	1,285	<0.001	**-0.16**	-0.23	-0.09	1184	**-0.20**	-0.30	-0.10	479
Within 30 days	**0.04**	-0.09	0.18	1,285	0.001	**0.00**	-0.14	0.15	1184	**-0.05**	-0.24	0.15	479
Within 90 days	**0.40**	0.12	0.69	1,290	0.063	**0.32**	0.01	0.63	1190	**0.33**	-0.09	0.75	476
Within 180 days	**0.42**	0.13	0.71	1,271	0.063	**0.35**	0.04	0.65	1170	**0.26**	-0.16	0.68	471
Outpatient specialist													
Within 10 days	**0.32**	0.28	0.37	1,285	0.017	**0.32**	0.28	0.37	1184	**0.37**	0.31	0.42	479
Within 30 days	**0.39**	0.32	0.47	1,285	0.115	**0.38**	0.31	0.46	1184	**0.46**	0.36	0.56	479
Within 90 days	**0.65**	0.50	0.81	1,290	0.674	**0.62**	0.46	0.79	1190	**0.63**	0.40	0.86	476
Within 180 days	**0.80**	0.55	1.06	1,271	0.434	**0.75**	0.48	1.02	1170	**0.88**	0.49	1.27	471
Days in hospital													
Within 10 days	**3.30**	3.13	3.47	1,285	<0.001	**3.15**	2.97	3.32	1184	**3.31**	3.08	3.55	479
Within 30 days	**3.68**	3.38	3.99	1,285	<0.001	**3.42**	3.10	3.75	1184	**3.72**	3.28	4.16	479
Within 90 days	**4.01**	3.48	4.54	1,290	<0.001	**3.62**	3.08	4.17	1190	**4.09**	3.36	4.83	476
Within 180 days	**4.13**	3.44	4.82	1,271	<0.001	**3.67**	2.95	4.38	1170	**4.24**	3.27	5.21	471
Hazard ratio for death	**Cox regression analyses**					**Poisson analyses**				**Cox regression analyses**			
Within 10 days	**0.53**	0.31	0.91			**0.44**	0.24	0.80		**0.35**	0.17	0.73	
Within 30 days	**0.54**	0.36	0.82			**0.45**	0.28	3.06		**0.47**	0.26	0.83	
Within 90 days	**0.62**	0.45	0.86			**0.54**	0.38	0.77		**0.55**	0.35	0.86	
Within 180 days	**0.72**	0.54	0.97			**0.64**	0.47	0.88		**0.67**	0.45	0.99	

* Instrument 1 *'physician referral preference'* is defined as the share of the physicians' out-of-hours contacts who were referred in the period, for the opposite sex. F-statistics estimates the strength of the instrument. Adjusted for patient age, age squared, month, and weekday.
** Instrument 1 *'physician referral preference'* is defined as the share of the physicians' out-of-hours contacts who were referred in the period, for the opposite sex. F-statistics estimates the strength of the instrument. Adjusted for referral diagnosis group (ICPC-2) patient age, age squared, month, and weekday.
*** Instrument 2 *'physician referral preference 2'* is defined as the share of the physicians' out-of-hours contacts in the study period, excluding contacts from the same year, and the years before and after each index contact. F-statistics estimates the strength of the instrument. Adjusted for patient sex, age, age squared, month, and weekday.

**** Test of difference in estimates between multivariable adjusted and instrumental variable analyses (Hausman).

Table S4 Descriptives: Mean use of health services and risk of death in the study population, during the study period 2008-2016

	Days with PCP contacts			Days with outpatient contacts			Days in hospital			Risk of death		
Follow-up	**Mean (SD)**	Referred	Not referred	**Mean (SD)**	Referred	Not referred	**Mean (SD)**	Referred	Not referred	**Proportion**	Referred	Not referre
0-10 days	**1.9 (1.1)**	1.6 (0.9)	1.9 (1.1)	**0.2 (0.6)**	0.4 (0.8)	0.2 (0.5)	**1.6 (3.0)**	4.7 (3.6)	0.2 (1.7)	0.02	0.03	0.02
0-30 days	**3.9 (2.0)**	2.9 (2.0)	3.1 (2.1)	**0.5 (1.2)**	0.8 (1.4)	0.4 (1.0)	**2.4 (5.0)**	6.3 (6.5)	1.1 (3.5)	0.04	0.05	0.03
0-90 days	**5.7 (4.3)**	5.9 (4.4)	5.7 (4.3)	**1.2 (2.6)**	1.7 (2.9)	1.1 (2.4)	**3.7 (7.8)**	8.2 (10.2)	2.1 (6.2)	0.06	0.10	0.05
0-180 days	**5.8 (4.3)**	5.9 (4.4)	5.7 (4.3)	**2.1 (4.2)**	2.8 (4.8)	1.9 (4.0)	**4.9 (10.2)**	9.9 (12.8)	3.2 (8.4)	0.10	0.13	0.08

1. The Norwegian Directorate of Health. Control and Payment of Health Reimbursement Register Data base 2019 (KUHR-databasen 2019) 2019 [Available from: https://www.helsedirektoratet.no/tema/statistikk-registre-og-rapporter/helsedata-og-helseregistre/kuhr.
2. Hughes LD, McMurdo ME, Guthrie B. Guidelines for people not for diseases: the challenges of applying UK clinical guidelines to people with multimorbidity. Age Ageing. 2013;42(1):62-9.